

## Datenkompetenz AT für ein zukunftsfähiges Jurastudium

### Warum ein Grundverständnis von Data Mining auch Juristen angeht

Philipp Kürth, Student, Humboldt Universität Berlin

28. November 2018

LR 2018, Seiten 211 bis 217 (insgesamt 7 Seiten)

---

Werdende Rechtswissenschaftler finden sich beim Absolvieren der juristischen Staatsprüfung seit eineinhalb Dekaden formell sowie materiell in einer unveränderten Situation wieder. Die Inhalte von Lehre und Prüfungstoff richten sich ausschließlich nach den grundlegenden Kodifikationen unserer Rechtsordnung. Auch mit Stift, Klausurenblock und Gesetzestext aus Papier schreibt man die Prüfungen noch. Angesichts eines weiblichen Roboters genannt Sophia, welcher in einem TV-Interview schlagfertig auf eine Frage nach seinem Fortpflanzungsbegehren eingeht, ist der auf die juristische Prüfung folgende Berufseinstieg ein Aufbruch in nicht-kartographierte Gewässer ohne Kompass oder Sextant.

1

#### **I. Die Handhabbarkeit rechtlicher Probleme der Technologisierung knüpft an das Verständnis der technischen Hintergründe an**

Dabei verändert die vielfältige Zunahme an datenerfassenden Sammelstellen die Interaktion von natürlichen und juristischen Personen signifikant und verursacht eine Verschiebung von Transparenz und Handlungsmöglichkeiten. Um die Einfallstore rechtlicher Probleme in der massenhaften Datensammlung, -verarbeitung und automatisierten Entscheidungsfindung zu erkennen und schließen zu können, gilt es, die Aussagekraft weniger Datenpunkte und den Prozess der Datenverarbeitung zu verstehen.

2

Maschinen, deren kognitive Kompetenzen (oder zumindest die Simulation einer solchen) über die ihrer menschlichen Schöpfer hinauswächst, sind beständig Gegenstand von Verheißungen über Arbeitsmarktentwicklungen, technologischen Fortschritt und den Gestaltwechsel gesellschaftlicher Prozesse und Interaktion. Ob und wie schnell dieser Vorgang eine dem Menschen überlegene Spezies hervorbringt, wird unterschiedlich beurteilt. Die Frage nach einer absehbaren totalen Ablösung des menschlichen Individuums im Wertschöpfungsprozess soll jedoch nicht Gegenstand dieser Erörterung sein. Auch ob die dahin tendierende Prognose ebenfalls die Sonderausprägung der menschlichen Spezies „Juristen“ betrifft, oder ob für diese Subjekte ein

3

Ausnahmetatbestand gilt, wurde bereits von vielen anderen kundigen Publizisten und Wissenschaftlern tiefgreifend dezidiert und soll daher außenvor bleiben. Im Mittelpunkt steht nachfolgend der Entwicklungsprozess, der diesem Szenario vorgelagert ist, und die Fragestellung, wie werdende Juristen befähigt werden, an diesem Prozess gestaltend mitzuwirken.<sup>1</sup> Denn jedenfalls werden wir Menschen, bis es zu unserer Überflüssigkeit gekommen ist, in unseren problembezogenen kognitiven Fähigkeiten und notwendigerweise in unseren Erwerbstätigkeiten bald Werkzeuge oder auch Mitstreiter an unserer Seite finden, die unsere Handlungsstrukturen generalüberholen, uns ein gewisses Tätigkeitsportfolio abnehmen und dafür neue Aufgaben zuweisen werden. Markant an diesem evolutionären Hergang ist, dass wir aufgrund der technischen Entwicklung in neuen Dimensionen und Prozessen denken müssen. Auch angehende Juristen müssen diesen Fortlauf verstehen, um als Funktionsträger im Markt zu bestehen.

## II. Datenkompetenz als Wissensbasis im Umgang mit neuen Rechtsfragen

Die Grundlage für das erforderliche Verständnis ist Datenkompetenz – oder wenn man so will das Verständnis von: 4

1. den Potenzialen dessen, was die „Raffination“<sup>2</sup> nur weniger Kennzahlen uns an tiefgreifenden Einsichten gestattet und
2. den damit verbundenen Folgen.

Denn Daten sind der Treiber der Möglichkeiten, Algorithmen „lernen“ und mittels statistischer Auswertungen schlauer als Menschen werden zu lassen. Daten sind der Rohstoff, aus dem Maschinensoftware ihre kognitiven Kompetenzen zieht. Der Wert von Daten im 21. Jahrhundert zeichnet sich letztendlich durch den Umstand ab, dass Daten für digitale Unternehmen den Charakter einer Ersatz-Vergütung angenommen haben, die bei unentgeltlichen Dienstleistungen an die Stelle einer finanziellen Gegenleistung tritt. Ihr Wert liegt darin, dass sie es erlauben, Schlüsse zu ziehen, dass sie Erkenntnis hervorbringen und (wenn auch meist nur einseitig) Transparenz schaffen. Eine zukünftig wahrscheinlich entstehende Interdisziplinarität von Rechts- und Datenwissenschaften, findet in der juristischen Ausbildung bislang wenig Würdigung. Und so mangelt es an der Vermittlung des Grundverständnisses von der nachfolgend illustrierten Wesensart des Rohstoffs Daten. 5

Bereits wenige Daten ermöglichen präzise Angaben über eine Person, ihren Kontext, ihr Verhalten und ihre Vorlieben. Das verdeutlicht die Gefahr, die sich aus der Offenlegung einfacher demographischer Informationen ergibt und die L. Sweeney<sup>3</sup> bereits 1990 6

<sup>1</sup> Zur Thematik auch: Tianyu Yuan, Weiter lernen ohne Zukunft. Legal Tribune Online, 25. Oktober 2018.

<sup>2</sup> Ein Terminus, den Andreas Weigend in seinem Werk „Data for the people“ für die *Verarbeitung der Rohdaten zu aussagekräftigen Analysen* gebraucht

<sup>3</sup> Latanya Sweeney, Uniqueness of Simple Demographics in the U.S. Population. Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-WP4. Pittsburgh, 2000.

feststellte und Philippe Golle vom Palo Alto Research Center bestätigte.<sup>4</sup> Gollés Studie zeigt, dass im Jahr 2000 63% der Bürgerinnen und Bürger der USA eindeutig identifizierbar waren allein anhand der Zensusdaten Geschlecht, PLZ und Geburtsdatum. Die Anonymität des Einzelnen – genauer gesagt die Größe der Gruppe, mit denen der Einzelne identische Daten teilt – ist abhängig von ihrem Alter sowie der Bevölkerungsdichte ihres Bezirks. Sachkenntnis über die Aussagekraft einer Kombination von nur wenigen Kennzahlen ist der Grundstein zur rechtlichen Beurteilung von Chancen und Risiken der Datenerhebung, -speicherung und -weiterverarbeitung. Das erforderliche Grundverständnis ist indes bei der Allgemeinheit nur schwach ausgeprägt.

Aber nicht nur lässt sich unsere Identität anhand weniger Datenpunkte feststellen, auch lässt sich an Spuren, die eine Person kontrolliert in den Datensammelstellen der sozialen Plattformen hinterlässt, akkurat ablesen, welche Persönlichkeitsmerkmale sie charakterisieren. Yoram Bachrach, Michal Kosinski u.a.<sup>5</sup> zeigen, dass sich aus der Anzahl an Facebook-Freunden und Gruppen, in denen ein Facebook-Nutzer Mitglied ist, sowie aus der Frequenz der geteilten Bilder und Likes, die dieser hinterlässt, in hohem Maße mit seinen Persönlichkeitsmerkmalen korrelieren. Auch lässt sich aus diesen Kennzahlen akkurat die Ausprägung etwa der Merkmale Extraversion und Neurotizismus ablesen.

„Data-Mining automatisiert den Prozess, nützliche Muster zu entdecken, um Regelmäßigkeiten aufzuspüren, die eine anschließende Entscheidungsfindung ermöglichen.“<sup>6</sup> Die kunstgerechte Verwertung von Rohdaten durch die „systematische Anwendung statistischer Methoden auf die Datenbestände“<sup>7</sup> ist ein Tool zur Desillusion. Die Möglichkeiten und Gefahren, die aus der geschickten Kombination der richtigen Datenpunkte hervorgehen und die Tatsache zu verstehen, dass sich aufgrund der massiven Erfassung von Daten im 21. Jahrhundert für die Beurteilung nahezu aller Charakteristika einer Person Kennzahlen aus Datenbanken heranziehen lassen, ist der Grundstein, um rechtliche Risiken erkennen und jurieren zu können.

### III. Einfallstore für Benachteiligung im Datenverarbeitungsprozess

Neben einem abstrakten Grundverständnis von Data-Mining sollte jedoch auch der Bezug von Datenverarbeitung zur Kerntätigkeit der Juristen selbst in die juristische Grundlagenausbildung des 21. Jahrhunderts einfließen. Die zum Massengeschäft und Profittreiber gewordene Verarbeitung des Rohstoffs Daten bringt nämlich gleichwohl rechtliche Problemfelder hervor. Diese gewinnen an Gewicht, wenn die Daten als Input von Maschinen dienen, die mit ihnen zu einer dem Menschen kognitiv überlegenen Art trainiert werden sollen. Die Risikopotenziale beim Umgang mit Daten zu kennen und

7

<sup>4</sup> Philippe Golle, Revisiting the Uniqueness of simple Demographics in the US Population. WPES'06, 30. Oktober 2006, Alexandria, VA, USA.

<sup>5</sup> Yoram Bachrach, Michal Kosinski, Thore Graepel u.a., Personality and Patterns of Facebook Usage. Web Science'12, 22.-24. Juni 2012, Evanston, IL, USA.

<sup>6</sup> Solon Barocas, Andrew D. Selbst, Big Data's Disparate Impact. California Law Review 104 (3), Juni 2006, Berkeley, CA, USA. S.676

<sup>7</sup> Wikipedia-Definition für den Anglizismus „Data-Mining“

regulieren zu können, wird zu einer der Kernkompetenzen von Juristen werden, während Prozesse der Datenerhebung und -analyse Einzug in jeden Prozess wirtschaftlicher Wertschöpfung halten.

Ein umfassend diskutierter Risikoschwerpunkt in der Datenverarbeitung ist das Problem der Diskriminierung durch Algorithmen, wie es Salon Barocas und Andrew D. Selbst beschreiben.<sup>8</sup> Basis für die rechtliche Abhilfe dieses Problems ist die Kenntnis über den Datenverarbeitungsprozess. Dieser lässt sich unterteilen in die Schritte (1.) Definition der Zielvariablen<sup>9</sup>, (2.) Erfassung und Kennzeichnung der Trainingsdaten, (3.) Auswahl der bei der Analyse beobachteten Attribute, die in die Auswertung einfließen und (4.) Erkenntnisgewinn auf Basis des resultierenden Modells. Auf jeder dieser Stufen kann eine von drei Quellen der Diskriminierung das Ergebnis der Datenauswertung indoktrinieren. Es können in den Daten existierende Muster von Diskriminierung reproduziert, Vorurteile von Entscheidungsträgern eingearbeitet oder in der Gesellschaft auftretende Benachteiligungen reflektiert werden.

8

Etwa die im ersten Schritt des Verarbeitungsprozesses stattfindende subjektive Übersetzung eines Problems in formale Bedingungen, mit denen ein Algorithmus arbeiten kann<sup>10</sup>, stellt ein solches Einfallstor für Diskriminierung dar. Im zweiten Schritt des Verarbeitungsprozesses müssen die auszuwertenden Daten mit einem Label versehen werden, um dem zu trainierenden Algorithmus die Verknüpfung aufzuzeigen, die er nach dem Trainingsprozess zwischen unbekanntem Datenmaterial selbst herstellen soll. Diese Klassifizierung der auszuwertenden Daten im zweiten Schritt bestimmt, was ein Algorithmus im Data-Mining-Prozess finden wird. Wird durch das Labeling eine Personengruppe benachteiligt, wird sich dies bei der Datenverarbeitung durch den Algorithmus fortsetzen, da er die Diskriminierung „antrainiert“ bekommen hat. Die Gefahr, für die der dritte Handlungsschritt des Data-Minings empfänglich ist, ist die Über- oder Unterrepräsentation von bestimmten Datenquellen (wie klassischerweise Personengruppen) hinsichtlich der für die Auswertung herangezogenen Attribute. „Die Qualität und Repräsentativität der Aufzeichnungen kann in einer Weise variieren, die mit der Klassenzugehörigkeit korreliert.“<sup>11</sup> Die Gefahr, die sich aus dem vierten Schritt der Datenraffination und der anschließenden Entscheidungsfindung ergibt, betrifft die darauffolgende Automatisierung des gesamten Entscheidungsprozesses. Diese könnte die zuvor in den Datenverarbeitungsprozess implizierte „Voreingenommenheit gegenüber Personen, die an früheren Entscheidungen beteiligt waren, in eine formalisierte Regel verwandeln, die die Perspektiven aller zukünftigen Beteiligten systematisch verändern würde.“<sup>12</sup> Alle vier Handlungsschritte können dazu führen, dass

9

<sup>8</sup> Solon Barocas, Andrew D. Selbst, Big Data's Disparate Impact. California Law Review 104 (3), Juni 2006, Berkeley, CA, USA. S.671-732.

<sup>9</sup> Der veränderliche Wert, der als Erkenntnisgegenstand das Ergebnis des Data-Mining-Prozesses darstellt

<sup>10</sup> Diese beinhaltet konkret die Umformulierung des Problems in die Werte, die die Zielvariable annehmen kann

<sup>11</sup> Solon Barocas, Andrew D. Selbst, Big Data's Disparate Impact. California Law Review 104 (3), Juni 2006, Berkeley, CA, USA. S.684.

<sup>12</sup> Ebenda, S.682

die später erfassten Werte eine Gruppe benachteiligen. Davon separat zu betrachtende Probleme sind die mangelnde Durchsichtigkeit des Analyseprozesses und die damit korrelierend schwere Verortung einer auftretenden Diskriminierung in einem der Prozessschritte.

#### IV. Ein neuer Eigentumsbegriff in Bezug auf Daten?

Zum allgemeinen Teil der Datenkompetenz zählt auch das Wissen um die beschränkte Kontrolle über einzelne Datenpunkte sowie die noch zu präzisierenden Rechte des Einzelnen an den ihm zuzuordnenden Daten. Die beschränkte Handlungsmacht wird beispielsweise dadurch begründet, dass der Zeitaufwand einer Prüfung der Informationen, welche der im vorigen Abschnitt geschilderten Verarbeitung zugrunde liegen, im Verhältnis zum Nutzen gegebenenfalls nicht zu rechtfertigen wäre. Aufgrund der Menge existierender Daten und der Frequenz ihrer fortwährenden Neuerfassung ist eine Garantie, dass alle Daten korrekt sind, nicht möglich. Für eine dem Einzelnen zugutekommende Verwendung der Daten durch die Raffinationen muss der Fokus auf die Fähigkeit gesetzt werden, Daten ergänzen zu können, also neue Daten mit vorhandenen zu verknüpfen.<sup>13</sup> Signifikanter noch ist die Tatsache, dass diejenigen Daten, die einem Individuum zuzuordnen sind, mitunter nicht von diesem selbst, sondern von anderen hinterlassen wurden. Die Identität einer Person lässt sich auch nicht ausschließlich durch die Daten bestimmen, die sie selbst über sich preisgibt. Ebenso aussagekräftig ist der mit ihr verbundene „soziale Graph“ – also die zwischenmenschlichen Beziehungen und die Kommunikation mit anderen, die sich in sozialen Netzwerken abbilden.<sup>14</sup> Es liegt folglich nicht allein in der Hand des Einzelnen, was über ihn kundgegeben wird und ob die gezogenen Schlüsse auch auf unverfälschten Daten beruhen.

10

Daraus folgt ein wichtiger Schluss für das Verständnis von Rechten an Daten: Unsere Daten gehören uns nicht allein, sie sind vielmehr mit Daten über andere verwoben<sup>15</sup>; „Daten können das Eigentum von mehr als einem [Eigentümer] gleichzeitig sein. Tatsächlich geht es beim Eigentum an Daten nicht darum, die alleinige Entscheidungsgewalt über die Bestimmung der Daten zu besitzen [...]. Stattdessen bedeutet das Eigentum an Daten die Fähigkeit, auf sie zuzugreifen und potenziell zu nutzen.“<sup>16</sup> Ein Recht zur Entfernung einzelner Daten wäre hingegen ein weniger geeignetes Instrument zur Ermöglichung von Einflussnahme des Einzelnen auf die ihm zuzuordnenden Daten. Sie aus ihrem Kontext zu entfernen, verändert nämlich auch den Kontext der Daten über andere, indem die mit den gelöschten in Verbindung stehenden Daten allein zurückbleiben.<sup>17</sup>

11

<sup>13</sup> So der Vorschlag von Andreas Weigend in: Data for the People. Murmann Publishers GmbH, Hamburg 2017. S.234.

<sup>14</sup> Andreas Weigend, Data for the People. Murmann Publishers GmbH, Hamburg 2017. S.104f.

<sup>15</sup> Ebenda, S.198

<sup>16</sup> Ebenda, S.202

<sup>17</sup> Ebenda, S.253

## V. Handlungsfähigkeit und Einflussnahme des Individuums auf ihm zugehörige Daten

Trotz der dargestellten Tatsachen muss der Einzelne in der Lage sein, „freie Entscheidungen auf Grundlage der eigenen, durch Datenfirmen ermittelten Präferenzen und Muster zu treffen“<sup>18</sup> – also handlungsfähig gegenüber den Konzernen, deren Geschäftsmodell die Erfassung und Raffination von Daten ist, zu sein. Den Einfluss dieser Datenportale, aus denen wir unsere täglichen Informationen ziehen und in die wir im Gegenzug Informationen unseres Alltags, unsere Vorlieben und unser Verhalten einspeisen, zeigt ein Experiment zu sozialem Einfluss und politischer Mobilisierung, durchgeführt während der Halbzeitwahlen 2010.<sup>19</sup> Ausgangspunkt war die Grundannahme, dass sich politisches Verhalten durch soziale Netzwerke verteilt. Bestätigt wurde diese Hypothese dahingehend, dass Facebook-Nutzer, welche in ihrem Newsfeed die Profilbilder der Freunde angezeigt bekamen, welche bereits gewählt hatten, eher auch selbst an der Wahl teilnahmen als diejenigen Nutzer, welche keine Information erhielten oder nur die Anzahl der Freunde sehen konnten, die von sich angaben, gewählt zu haben.<sup>20</sup> Die Statistiker untersuchten auch, wie sich das Wahlverhalten und die Informationsweitergabe darüber auf Facebook auf die engen Freunde auswirkte, die das soziale Netzwerk nutzten. Sie fanden heraus, dass sich die Wahrscheinlichkeit, mit der ein Nutzer wählen würde, um 0,224% erhöhte für jeden engen Freund, der in seinem Newsfeed die Bilder seiner bereits wahlbeteiligten Freunde einsehen konnte.<sup>21</sup> Im Ergebnis war die Verbreitung der Informationen über die an der Wahl teilnehmenden Freunde für 340.000 Wahlbeteiligte mehr verantwortlich.<sup>22</sup> Die dargestellte Analyse mag als solche nicht weiter bedenklich erscheinen; ist doch eine hohe Wahlbeteiligung wünschenswert. Zu bedenken ist jedoch, dass den Facebook-Nutzern nicht klar war, dass ihre Entscheidung wählen zu gehen, vom Inhalt ihres Newsfeeds abhing. Im Allgemeinen arbeitet der Algorithmus von Facebook auf Basis des Interaktionsverhaltens eines Nutzers heraus, welche Inhalte seinen Interessen entsprechen und bespielt den Newsfeed damit. Wenn die Empfindung des Einzelnen über seine Vorlieben sich jedoch „von der Interpretation unterscheidet, zu der ein Algorithmus gelangt ist“<sup>23</sup>, muss das Individuum dennoch selbst die mit seinen Präferenzen im Zusammenhang stehende Entscheidung auf Grundlage einer eigenen Bewertung treffen können.<sup>24</sup> Sich die Entscheidung nicht ebenfalls von einem Algorithmus abnehmen zu lassen, ist die Basis für selbstbestimmtes Handeln.

12

---

<sup>18</sup> Ebenda, S.24

<sup>19</sup> Robert M. Bond, Christopher J. Fariss, Jason J. Jones u.a., A 61-million-person experiment in social influence and political mobilization. *Nature* 489, 13. September 2012. S.295-298.

<sup>20</sup> Ebenda, S.296

<sup>21</sup> Ebenda

<sup>22</sup> Robert M. Bond, Christopher J. Fariss, Jason J. Jones u.a., A 61-million-person experiment in social influence and political mobilization. *Nature* 489, 13. September 2012. S.297

<sup>23</sup> Andreas Weigend, *Data for the People*. Murmann Publishers GmbH, Hamburg 2017. S.178

<sup>24</sup> Vgl. ebenda, S.258

Selbstbestimmung ist der Grundstein für ein freiheitlich-demokratisches Zusammenleben von Individuen. Diese Selbstbestimmung scheint durch die technischen Entwicklungen des 21. Jahrhunderts in einer neuen Art und Weise herausgefordert zu werden. Die Rechtswissenschaftler der nächsten Generation sollten sich das Schicksal gesellschaftlicher Grundwerte nicht durch technische Feinheiten aus der Hand nehmen lassen müssen. Dazu müssen sie die Grundzüge der Entwicklungen nachvollziehen können.

13